

Faculty of Engineering and Information Technology
University of Technology, Sydney

An Agent-Based Service Oriented Architecture for Risk Mining

A thesis submitted in partial fulfilment of

the requirements for the degree of

Master of Science in Computer Science (research)

By

Jiahang Chen

July 2012

Certificate

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Jiahang Chen

Acknowledgment

Foremost, I would like to express the deepest appreciation to my supervisor, Professor Longbing Cao, for his professional guidance, persistent help and continuous support throughout my Master study and research.

I would like to thank Dr. Xinhua Zhu for his patient guidance, and scientific advice. Without his generous support this dissertation would not have been possible.

Besides, I offer my regards and blessings to all of my co-workers at lab, and thank them for their support in my research and during the completion of this dissertation.

Last but not the least, I would like to thank my family: my parents, Without their encouragement, finishing this dissertation would be impossible; without them, nothing would have any value.

Jiahang chen

July 2012 @ UTS

Table of Contents

Certificate.....	I
Acknowledgment	II
Abstract	4
1. Introduction.....	6
1.1 Background	7
1.2 Motivation	9
1.3 The Challenge.....	12
1.4 The Goals	13
1.5 The Contribution	14
1.6 The Thesis Organization	16
2. State of the Art	19
2.1 Agent-based distributed data mining.....	20
2.1.1. The requirement of agent-driven distributed data mining.....	20
2.1.2. The requirement of data mining-driven agent.....	22
2.1.3. The challenge for the integration and interaction	22
2.1.4. Extendible multi-agent data mining framework	23
2.2 Service oriented architecture	27
2.2.1. Elements of service-oriented architecture	27
2.2.2. Characters of service-oriented architecture.....	29
2.2.3. Three styles of SOA	29
2.3 Ensemble learning	30

3. Agent-Based Service Oriented Architecture	34
3.1. The integration methodology	35
3.2. Overview of the architecture	37
3.3. Distributed service bus	40
3.4. Pluggable algorithm	45
3.5. Conclusion.....	50
4. Agent-Based Business Process Management	51
4.1. Introduction	52
4.2. Business Process Engine	53
4.3. Integrating workflow into risk mining	56
4.3.1. Risk mining model	57
4.3.2. Descriptive workflow evaluation	57
4.3.3. Example.....	60
4.3.4. Evaluation	64
5. Agent-Based Ensemble Learning Strategies.....	65
5.1. Introduction	66
5.2. Ensemble methods for risk mining.....	67
5.2.1. Under-sampling based bagging for imbalance learning.....	67
5.2.2. Cost-sensitive learning with adaptive boosting	68
5.3. Performance measurement	70
6. Case Studies	72
6.1. Online banking risk management system.....	73
6.1.1. Background	73
6.1.2. Solution	75

6.1.3. Outcome and evaluation.....	78
6.2. Student risk management system	82
6.2.1. Background	82
6.2.2. Solution	83
6.2.3. Outcome	87
7. Conclusion	94
7.1. Summary	94
7.2. Future work	96
Appendix A: List of Publications.....	97
Bibliography.....	98

Abstract

Risk Mining (RM) is the process of analyzing data including risk information by data mining methods, with the mining results for risk prevention. In the last few years, some researchers have proposed the combination of data mining and agent technology (agent mining) to improve the performance of data mining methodology in the heterogeneous business environments. However, problems exist for further research with the application of risk mining systems in real industry environments to enhance the robustness of system architect, dynamic business process and model accuracy etc.

Therefore, in this thesis we present an Agent-based Service-oriented Risk Mining Architecture (ABSORM), which has been designed to facilitate the development of agent mining systems to address the above issues. This thesis focuses on developing the following strategies:

- **The integration of agent technology with web service.**

In this framework, we propose a new and easier method, by which the system functions are not integrated into the structure of the agents, rather modeled as distributed services and applications which are invoked by the agents acting as controllers and coordinators. Therefore, techniques developed in this framework can improve the interoperability between different modules, distribution of resources, and the lack of dependency of programming languages.

- **The integration of agent technology with business process management.**

In this work, we develop the autonomous agents that can collaborate in a business flow, which not only increases the reusability of the system, but also eases the system development in terms of re-usability of the computational resources. A group of agents solves problems in the following way: each individual agent solves

the problem individually, and then interacts with each other to finalize a business process.

- **The integration of agent technology with ensemble learning methods.**

In this thesis, we are interested in developing agent-based ensemble learning strategies for risk mining: each ensemble agent individually gathers the evidence about model evaluation, and then ensembles learning methods like bagging and boosting is used to obtain prediction from the individually gathered evidence. Agent based ensemble learning can provide a critical boost to risk mining where predictive accuracy is more vital than model interpretability.

The proposed architecture has been evaluated for building an online banking fraud detection system and a student risk management system. These two applications have been proved to be a sophisticated, yet user friendly, risk analysis and management tool. They are modular, interactive, dynamic and globally oriented.

Keywords: risk analysis and management, risk mining, service oriented architecture, multi-agent system, BPEL, ensemble learning